Efficiency Evaluation When Modelling Nairobi Security Exchange Data Using Bilinear and Bilinear-Garch (Bl-Garch) Models

ADOLPHUS WAGALA
Chuka University College
Department of Business Administration
P.O BOX 109-CHUKA-60400
Kenya

ALI S. ISLAM
Egerton University
Department of Mathematics

DANKIT K. NASSIUMA
Kabarak University
School of Engineering

Abstract

In this paper, the weekly returns of the Nairobi Securities Market (NSE) are modelled using bilinear models and the bilinear-GARCH models so as to determine the most efficient and adequate model for forecasting of the Nairobi Equity market. The data used was obtained from the Nairobi Stock Exchange (NSE) for the period between 3rd June 1996 to 31st 30th October 2011for the company share prices while for the NSE 20-share index was for period between 2nd March 1998 to 30th October 2011.The share prices for three companies; Bamburi Cement, National Bank of Kenya and Kenya Airways which were selected at random from each of the three main sectors as categorized in the Nairobi Stock Exchange were used. The results indicate that the combination of bilinear-GARCH model is more adequate and efficient in modelling the weekly returns of the Nairobi Securities Exchange.

Keywords: Bilinear, Efficiency, GARCH, Bilinear-GARCH

1.0 Introduction

Stock market volatility is one of the widely studied aspect of the equity markets world over mainly by use of the Autoregressive Moving Average (ARMA) models with Autoregressive Heteroscedasticy (ARCH) errors. The focus has been on the ARCH family of models thereby ignoring the other suitable non linear models like the bilinear model proposed by Granger and Anderson (1978). In the bilinear model, the unconditional moment structure is similar to that of ARCH that it can be mistaken for ARCH (Bera and Higgins, 1997 Weiss, 1986). The bilinear model captures the nonlinear dynamics in the conditional mean while the ARCH models captures the nonlinear dynamics in the conditional variance. As such, the two models can be treated as 'competitors' since they have similar statistical properties or as 'complementary' models since the bilinear model captures the nonlinear dependence in the conditional mean while the ARCH model captures the nonlinearity due to the variance. The outline of this paper is as follows. In section 2, a review of the bilinear models are given. Section 3. presents the Autoregressive Conditional Heteroscedasticity (ARCH) Models while the Methodology is given in section 4. The Results and Discussions are presented in section 5 while section 6 is a list of References.

2.0 Bilinear Models (B.L.)

Following Granger and Anderson (1978) , Subba (1981), Subba and Gabr (1984), a time series $\{X_t\}$ is said to follow a bilinear time series denoted by BL(p,q,m,k) if it satisfies the equation

$$X_t + \sum_{i=1}^{p} a_i X_{t-i} = \sum_{j=0}^{q} c_j \varepsilon_{t-j} + \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij} X_{t-j}\varepsilon_{t-k} + \varepsilon_t \qquad 1$$

where $\{\varepsilon_t\}$ is a sequence of *i.i.d* random variable, usually but not always with zero mean and variance σ_ε^2 and $c_0=1$, a_i ,b_{ij} and c_j are model parameters. It is easy to see that bilinear model is a special case of ARMA (p,q) model.

Using lag (Backshift) operator, equation 1 can be specified as

$$\phi(B)w_t = (\lambda(B)\varepsilon_{t-1})\psi(B)w_{t-1} + \theta(B)\varepsilon_t \qquad 2$$

where

$$w_t = (1-B)^{d_1} X_t,\ \varepsilon_t \sim WN(0,\sigma_\varepsilon{}^2) \qquad 3$$

In equation 2, $\psi(B) = \psi_1 + \psi_2 B + \ldots + \psi_p B^{p-1}$ and $\lambda(B) = 1 + \lambda_1 B + \ldots + \lambda_r B^{r-1}$ .This is equivalent to the Subba (1981) BL(p,q,m,k) class of models (Davidson, 2008).

Bilinear models have been applied in geophysics data (Subba, 1988), Spanish economic data (Maravall, 1983) and in solar physics data by (Subba and Gabr, 1984). These models are particularly attractive in modelling processes with sample paths of occasional sharp spikes (Subba and Gabr, 1984). In the bilinear model although the conditional variance is constant, the conditional mean is augmented with the interaction terms between the past observations and the innovations. This may increase the predictability of the dependent variable (Bera and Higgins, 1997).

2.1 Estimation of parameters in bilinear models

Consider the BL(p,0,m,k) model of the form

$$X_t + a_1 X_{t-1} + \ldots. + a_p X_{t-p} = \alpha + \sum_{i=1}^{m}\sum_{j=1}^{k} b_{ij} X_{t-i}\varepsilon_{t-j} + \varepsilon_t \qquad 4$$

where $\{\varepsilon_t\} \sim (0,\sigma_\varepsilon^2)$. (Here the, MA terms have been dropped and a constant α has been added to the R.H.S to facilitate the fitting of such models to non mean corrected data).

The likelihood function of the unknown parameters is constructed, given N observations $X_1, X_2,..., X_N$. Since the model involves lagged values of the $\{X_t\}$, one can not evaluate the residuals for the initial stretch of data. The conditional likelihood based on $X_{\gamma+1}, X_{\gamma+2},..., X_N$, given $X_1, X_2,..., X_N$ where $\gamma = \max(p,m,k)$ is thus considered.

Let $\underline{\theta} = (\theta_1, \theta_2,..., \theta_n)$ denote the complete set of parameters $\{a_i\}$, $\{b_{ij}\}, \alpha$ i.e. set $\theta_i = a_i, i = 1,2,..., p$, $\theta_{p+1} = b_{11}, \theta_{p+2} = b_{12},..., \theta_{p+mk} = b_{mk}, \theta_{p+mk+1} = \alpha$ and write n=p+mk+1 to denote the total number of parameters.

The joint probability density function of $\varepsilon_\gamma, \varepsilon_{\gamma+1}, \varepsilon_{\gamma+2},..., \varepsilon_N$ is given by

$$\left(1/2\pi\sigma_\varepsilon^2\right)^{(N-\gamma)/2} \exp\{\frac{1}{2\sigma_\varepsilon^2}\sum_{t=\gamma+1}^{N}\varepsilon_t^2\} \qquad 5$$

and since the Jacobian of the transformation from $\{X_t\}$ to $\{\varepsilon_t\}$ is unity, equation 5 also represents the likelihood function of ***θ***, given $\{X_t; t = \gamma+1,..., N\}$. The (conditional) maximum likelihood estimates of $\theta_1, \theta_2,..., \theta_n$ are thus given by maximizing equation 5 or equivalently by minimizing

$$Q(\boldsymbol{\theta}) = \sum_{t=\gamma+1}^{N}\varepsilon_t^2 \qquad 6$$

The minimization is performed numerically: for a given set of values $(\theta_1, \theta_2,..., \theta_n)$ then $\{\varepsilon_t\}$ is evaluated recursively from equation 1 and then the Newton-Raphson method is used to minimize Q(**θ**) (Subba, 1981). The Newton-Raphson iterative equations for minimization of Q (**θ**) are given by,

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \mathbf{H}^{-1}(\boldsymbol{\theta}^{(i)})\mathbf{G}(\boldsymbol{\theta}^{(i)}) \qquad 7$$

where $\boldsymbol{\theta}^{(i)}$ is the vector of parameter estimates obtained at the i[th] iteration, and gradient vector **G** and Hessian matrix **H** are given respectively as,

$\left[\frac{\partial Q}{\partial\theta_1},\frac{\partial Q}{\partial\theta_2},\ldots,\frac{\partial Q}{\partial\theta_n}\right]'$ and $\mathbf{H}(\boldsymbol{\theta})=\left\{\frac{\partial^2\theta}{\partial\theta_i\partial\theta_j}\right\}$. The partial derivatives of Q with respect to $\{\boldsymbol{\theta_i}\}$ are shown to be $\frac{\partial^2 Q}{\partial\theta_i}=2\sum_{t=\gamma+1}^{N}\varepsilon_t\frac{\partial\varepsilon_t}{\partial\theta_i}$, i=1,2,…,n

$$\frac{\partial^2 Q}{\partial\theta_i\partial\theta_j}=2\sum_{t=\gamma+1}^{N}\frac{\partial\varepsilon_t}{\partial\theta_i}\frac{\partial\varepsilon_t}{\partial\theta_j}+2\sum_{t=\gamma+1}^{N}\frac{\partial^2\varepsilon_t}{\partial\theta_i\partial\theta_j},\ ij=1,2,\ldots,N \qquad 8$$

Subba and Gabr (1984) developed a neat set of recursive equations for those derivatives as follows. Differentiating equation 1 with respect to each of the parameters the following are obtained

$$\frac{\partial\varepsilon_t}{\partial a_i}+\phi(a_i)=X_{t-i}, \qquad i=1,2,\ldots,p$$

$$\frac{\partial\varepsilon_t}{\partial a_i}+\phi(b_{ij})=-X_{t-i}\varepsilon_{t-i}, \qquad i=1,2,\ldots,\mathrm{m} \qquad j=1,2,\ldots,k$$

$$\frac{\partial\varepsilon_t}{\partial a_i}+\phi(\alpha)=-1 \text{ ,where } \phi(\theta_1)=\sum_{i=1}^{m}\sum_{j=1}^{k}b_{ij}X_{t-i}\frac{\partial\varepsilon_{t-i}}{\partial\theta_i}$$

Assuming the initial conditions

$$\varepsilon_t=\frac{\partial\varepsilon_t}{\partial\theta_i}=0 \quad t=1,2,\ldots,\gamma \qquad i=1,2,\ldots,n$$

the second order derivatives satisfy

$$\frac{\partial^2\varepsilon_t}{\partial a_i\partial a_{i'}}=0,\ \frac{\partial^2\varepsilon_t}{\partial a_i\partial\alpha}=0,\ i,i'=1,\ldots,p$$

$$\frac{\partial^2\varepsilon_t}{\partial a_i\partial b_{j1}}+\psi(a_i,b_{j1})=-X_{t-j}\frac{\partial\varepsilon_{t-1}}{\partial a_i},\ \frac{\partial^2\varepsilon_t}{\partial b_{ij}\partial b_{i'j'}}+\psi(b_{ij},b_{i'j'})=-X_{t-i}\frac{\partial\varepsilon_{t-j}}{\partial b_{i'j'}}-X_{t-i'}\frac{\partial\varepsilon_{t-j'}}{\partial b_{ij}}$$

$$\frac{\partial^2\varepsilon_t}{\partial b_{ij}\partial\alpha}+\psi(b_{ij},\alpha)=-X_{t-j}\frac{\partial\varepsilon_{t-j}}{\partial\alpha},\ \frac{\partial^2\varepsilon_t}{\partial\alpha^2}=0$$

where $\psi(\theta_r,\theta_l)=\sum_{i=1}^{m}\sum_{j=1}^{k}b_{ij}X_{t-i}\frac{\partial\varepsilon_{t-i}}{\partial\theta_r\partial\theta_l}$.

For a given set of parameter values α ,{a_i}, {b_{ij}} the first and the second derivatives of Q can be evaluated from the above equations and hence the vector **G** and matrix **H** evaluated. The iteration equation 7 is now implemented.

When the final parameter estimate $\underline{\hat{\theta}}$ have been obtained, σ_e^2 is obtained as,

$$\hat{\sigma}_\varepsilon^2=\frac{1}{(N-\gamma)}Q(\hat{\theta})=\frac{1}{(N-\gamma)}\sum_{t=\gamma+1}^{N}\hat{\varepsilon}_t^2 \qquad 9$$

2.2 Least squares estimation of model parameters for bilinear models

Following the approach of Tong (1990), consider the bilinear model (equation 1). Rewriting it in a Markovian representation with monor changes in notation. Set p=max(p,q,m,k).

$$\xi_t=(\mathbf{A}+\mathbf{B}\varepsilon_t)\xi_{t-1}+\mathbf{c}\varepsilon_t+\mathbf{d}(\varepsilon_t^2-\sigma^2)$$

$$\mathbf{X}_t=\mathbf{H}\xi_{t-1}+\varepsilon_t \qquad 10$$

where ξ_t is a p-vector and

$$\mathbf{A}=\begin{bmatrix} a_1 & 1 & 0 & \cdots & 0 \\ a_2 & 0 & 1 & \cdots & 0 \\ \vdots & & & & 1 \\ a_p & 0 & 0 & \cdots & 0 \end{bmatrix} \qquad \mathbf{B}=\begin{bmatrix} b_1 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ b_p & 0 & \cdots & 0 \end{bmatrix}$$

$$\mathbf{c}=\begin{bmatrix} a_1+c_1 \\ \vdots \\ a_p+c_p \end{bmatrix} \qquad \mathbf{d}=\begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix} \qquad \mathbf{H}=\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

By convention, $c_j=0$ if $j>q$, $b_{jk}=0$ if $j>m$. Diagonality of the model implies the relation
B=dH. The converse is not true.

Following Guegan and Pham (1989), take $\boldsymbol{\theta}=(a_1,\ldots,a_p, c_1,\ldots,c_p, b_1,\ldots,b_p)$ as the fundamental parameter vector and assume that the representation in equation 18 is quasiminimal in the sense that there is no other Markovian representation with the same noise structure but with a state vector which is a linear transformation of the original state vector and has a smaller dimension. Further**,** assume that the model is invertible and stationary.

Let $(X_1,\ldots,X_N)$ denote the observations. It is plausible to estimate $\boldsymbol{\theta}$ by minimizing
$\sum_{t=1}^{N}\varepsilon_t^2(\tilde{\theta}/\xi_0)$ w.r.t $\tilde{\theta}\in\Theta_N$ where $\varepsilon_t(\tilde{\theta}/\xi_0)$ is the $\varepsilon_t(\theta/\xi_0)$ given by
$\varepsilon_t(\tilde{\theta}/\xi_0)=\mathbf{X}_t\text{-}\mathbf{H}\xi_{t-1}(\xi_0)$ with $\tilde{\theta}$ is the parameter vector. Intuitively, for this to make sense the effect of ξ_0 on $\varepsilon_t(\tilde{\theta}/\xi_0)$ should diminish as $t\to\infty$(see Tong,1990). Let there exist a stationary time series $\{\varepsilon_t(\tilde{\theta})\}$ such that $\varepsilon_t(\tilde{\theta}/\xi_0)\text{-}\varepsilon_t(\tilde{\theta})\to 0$ as $t\to\infty$. Note that $\varepsilon_t(\tilde{\theta})$ is then measurable w.r.t σ **-**algebra generated by X_s, $s\leq t$. Here, the model 18 is said to be invertible at $\tilde{\theta}$ relative to the observation $\{X_t\}$. A sufficient condition for this is $[E_\theta(\ln\left\|\tilde{A}-\tilde{c}\tilde{H}-d\tilde{H}X_t\right\|_\phi]<0$, where $\tilde{A}$, $\tilde{c}$ **,** $\tilde{d}$ and $\tilde{H}$ are given by $\boldsymbol{\theta}$.

A reasonable choice of Θ_N is suggested by the above sufficient condition and given as

$$\Theta_{N,\delta}=\{\tilde{\theta}\in\Theta_0:\prod_{t=1}\left\|\tilde{A}-\tilde{c}\tilde{H}-\tilde{d}\tilde{H}X_t\right\|_\phi<(1-\delta)^N\} \qquad 11$$

where Θ_0 is a given compact set and δ is a small positive number. The set Θ_0 is chosen large enough to include $\boldsymbol{\theta}$, the true parameter and the set of parameters satisfying the stationarity condition.

$$\text{Let } Q_N(\tilde{\theta})=\sum_{t=1}^{N}\varepsilon_t^2(\tilde{\theta}/\xi_0) \qquad 12$$

Let $\hat{\theta}_N$ be the minimizer of $Q_N(\tilde{\theta})$ over the set $\Theta_{N,\delta}$ **.** This is the LSE of $\boldsymbol{\theta}$ (Tong,1990).

2.3 Order selection for Bilinear models

Suitable values of p,m,k are determined by fitting a range of models covering various values of p,m,k and then selecting the model with the minimum value of the AIC defined as;

$$AIC=(N-\gamma)\log\hat{\sigma}_e^2+2 \text{ x number of fitted parameters (Akaike, 1977).}$$

Note that (N-γ) is the effective number of observations to which each model has been fitted. In using the AIC criterion, the goal is to strike a balance between reducing the magnitude of the residual variance and increasing the number of model parameters. This method requires that the upper bounds be set to p,m,k and then search for the various combinations within this bound.

This is clearly a nested search procedure. The Subba (1981) algorithm is used to accomplish this as follows;

1. For a given value of p, fit a linear AR(p) model.
2. Using the AR coefficients as initial values for the $\{a_i\}$, $\{b_{ij}\}$ and α, and setting initially b_{11}=0, fit a BL(p,0,1,1) model using the Newton-Raphson technique.
3. Fit BL(p,0,1,2) and BL(p,0,2,1) models using the parameters of the BL(p,0,1,1) model as the initial values and setting initially the remaining bilinear coefficients as zero.
4. Of the two models fitted in step 3, choose the one which has the smaller residual variance and use its parameters as starting values of fitting the BL (p, 0,2,2) model.
5. The procedure is continued until *m,k* have reached a common upper bound Γ. At each stage, a bilinear term of order (*m,k*) is fitted by considering bilinear terms of orders (*m-1,k*) , (*m,k-1*) , and choosing whichever model has residual variance to provide the starting values, with initial values for the remaining coefficients set to zero.
6. All previous steps are repeated for p=1,2,…,Γ. and the procedure terminates when the residual variance $\hat{\sigma}_e^2$ starts to increase as *m,k* increases. As a working rule, Γ should be at least as large as the order of the best AR model selected by the AIC criterion. The final choice of model is then made by selecting the model for which the AIC value is smallest (Priestly, 1980)

3.0 The Autoregressive Conditional Heteroscedasticity (ARCH) Models

The ARCH models were first introduced by Engle (1982) when modelling the United Kingdom inflation and provides a mechanism that includes past variances in the explanation of future variances (Engle, 2004). Following Wagala, Nassiuma, Ali and Mwangi (2012), an ARCH process can be defined in terms of the distribution of the errors of a dynamic linear regression model. The dependent variable y_t is assumed to be generated by

$$y_t = x_t'\xi + \varepsilon_t \quad t=1,\ldots,T \qquad 13$$

where x_t' is a kx1 vector of exogenous variables, which may include lagged values of the dependent variable and ξ is a kx1 vector of regression parameters. The ARCH model characterizes the distribution of the stochastic error ε_t conditional on the realized values of the set of variables $\psi_{t-1} = \{y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, ...\}$. The Engle's (1982) model assumes

$$\varepsilon_t / \psi_{t-1} \sim N(0, h_t) \qquad 14$$

where $h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \ldots + \alpha_q\varepsilon_{t-q}^2$ 15

with α_0 >0 and $\alpha_i \geq 0, i = 1,..., q$ to ensure that the conditional variance is positive.
An explicit generating equation for an ARCH process is

$$\varepsilon_t = \eta_t\sqrt{h_t} \qquad 16$$

where η_t ~*i.i.d* N (0,1) and h_t is given by equation 29. Since h_t is a function of ψ_{t-1} and is therefore fixed when conditioning on ψ_{t-1}, it is clear that ε_t as given in equation 30 will be conditionally normal with $E(\varepsilon_t / \psi_{t-1}) = \sqrt{h_t}E(\eta_t / \psi_{t-1}) = 0$ and $Var(\varepsilon_t / \psi_{t-1}) = h_t$, $Var(\eta_t / \psi_{t-1}) = h_t$. Hence the process (4) is identical to the ARCH process equation 14.

Despite the importance of the original ARCH model for many financial time series, a relatively long lag length in the variance equation with the problem of estimation of parameters subject to inequality restrictions is often called for to capture the long memory typical of financial data. To overcome problem of a relatively long lag, the Generalized ARCH (GARCH) model was developed by Bollerslev (1986) by proposing an extension of the conditional variance function which he termed as the generalized ARCH (GARCH) and suggested that conditional variance be specified as,

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \ldots + \alpha_q\varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \ldots + \beta_p h_{t-p} \qquad 17$$

with the inequality conditions $\alpha_0 > 0$, $\alpha_i \geq 0$ for *i=1,…,q,* $\beta_i \geq 0$ for *i= 1,…,p* to ensure that the conditional variance is strictly positive.

A GARCH process with orders p and q is denoted as GARCH (p, q) and this essentially generalizes the purely autoregressive ARCH to an autoregressive moving average model. The motivation for the GARCH process can be seen by expressing equation 17 as

$$h_t = \alpha_0 + \alpha(B)\varepsilon_t^2 + \beta(B)h_t \qquad 18$$

where $\alpha(B) = \alpha_1 B + \ldots + \alpha_q B^q$ and $\beta(B) = \beta_1 B + \ldots + \beta_q B^q$ are polynomials in the backshift operator B. Now, if the roots of $1 - \beta(Z)$ lie outside the unit circle, equation 18 be written as

$$h_t = \frac{\alpha_0}{1-\beta(1)} + \frac{\alpha(B)}{1-\beta(B)}\varepsilon_t^2 = \alpha_0^* + \sum_{i=1}^{\infty}\delta_i \varepsilon_{t-i}^2 \qquad 19$$

where $\alpha_0^* = \frac{\alpha_0}{[1-\beta(1)]}$ and the co-efficient δ_i is the co-efficient of B^i in the expression of $\alpha(B)[1-\beta(B)]^{-1}$.

The slope parameter β measures the combined marginal impacts of the lagged innovations while α, on the other hand captures the marginal impact of the most recent innovation in the conditional variance. When $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1$, then the process is weakly stationary and the conditional variance (σ_t^2) approaches the unconditional variance (σ^2) as time goes to infinity i.e $E(\sigma_{t+s}^2) \to \sigma^2$ as $s \to \infty$. However, when $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j > 1$ then the process is non stationary. When the parameter estimates in GARCH (p,q) models are close to the unit root but not less than unit, i.e $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j = 1$, for the GARCH process, the multi-step forecasts of the conditional variance do not approach the unconditional variance. These processes exhibit the persistence in variance/volatility whereby the current information remains important in forecasting the conditional variance. Engle and Bollerslev (1986) refer to these processes as the Integrated GARCH or IGARCH which does not possess a finite variance but are stationary in the strong sense (Nelson, 1990; Wagala *et al*, 2012).

The GARCH (p,q) process is effectively an infinite order ARCH with a rational lag structure imposed on the co-efficient. The GARCH process can parsimoniously represent a high-order ARCH-process (Bera and Higgins, 1993; Engle, 2004; Degiannakis and Xekalaki, 2004). GARCH(1,1) is often found to be the benchmark of financial time series modelling because such simplicity does not significantly affect the preciseness of the outcome.

Despite the empirical success, GARCH models have two major draw backs: First, they are unable to model asymmetry because in a GARCH model, positive and negative shocks of the same magnitude produce the same amount of volatility (i.e only the magnitude and not the sign of the lagged residuals determines the conditional variance). However, volatility tends to rise in response to “bad” news and fall in response to “good” news (Nelson, 1991). Also, in the GARCH models the non-negativity constraints imposed on the parameters are often violated by estimated parameters (Curto, 2002).

4.0 Methodology

4.1 The scope of the study

This study was focused on modelling the weekly NSE 20-share index and share prices for the three chosen companies namely, National Bank of Kenya Limited (NBK), Bamburi Cement and Kenya Airways from Nairobi Stock Exchange using the Bilinear (BL) and BL-GARCH models. The companies selected have been consistent in the NSE and are representative of the three sectors namely, Finance & Investment, Industrial & Allied and Commercial & Services categorized in the NSE. The two models chosen are able to capture the properties of financial data commonly referred to as the “stylized facts”.

4.2 Data collection

Secondary data was collected from the Nairobi Securities Exchange (NSE).

The NSE 20-share index was used in addition to the individual company share prices because the behaviour of the volatility of individual stocks has received far less attention in the literature when compared with studies on market indices. Furthermore, individual investors are more interested in the specific risk of the securities they hold rather than the market index; this justifies the need to study stock level data. Moreover, it has also been identified in the literature that basing an analysis on index data can lead to false perceptions of price change dependence, even when price changes of individual shares represented by the index are independent, because stocks which are not traded frequently affect the market index (Baudouhat, 2004). The three companies were randomly selected from the three sectors. These are the major sectors which are consistent and contribute a lot to the Nairobi stock market. The average weekly share prices for the following companies were used: National Bank of Kenya Limited (NBK), Bamburi Cement and Kenya Airways (KQ) for the period between 3rd June 1996 to 30Th October 2011. The NSE 20-share index was for period between 2nd March 1998 to 30th October 2011 was also modelled.

4.3 Data analysis

Time plots for the data were obtained in order to check its empirical characteristics of the data. The maximum likelihood estimation (MLE) assuming a normal distribution was utilized for modelling the data using the bilinear and bilinear-GARCH models. The models were diagnosed using the Log likelihood ratio test, AIC and the BIC. Model adequacy was carried out for all cases by examining the standardized residuals and squared residual correlations through Ljung-Pierce Q-statistics. The MSE was used to check on the efficiency of various models in addition to the residual plots.

5.0 Results and Discussions

5.1 Basic Analysis

In this study, four sets of data were used for modelling. They include the weekly average share prices for Bamburi Cement Ltd, National Bank of Kenya Limited (NBK), Kenya Airways (KQ) Ltd as well as the weekly average NSE 20 share index. The NSE 20-share index is a weighted mean with 1966 as the base year at 100. It was originally based on 17 companies and was calculated on a weekly basis. Bamburi Cement, Ltd. was founded in 1951 and manufactures cement in sub-Saharan Africa and Kenya. The Kenya airways' principal activities include passengers and cargo carriage. It was incorporated in 1977 as the East African Airways Corporation (EAA). The company was listed in the NSE in 1996 and has been a major player in the Nairobi stock market. The National Bank of Kenya Limited (NBK) was incorporated on 19th, June 1968 and officially opened on Thursday 14th, November 1968.

The series were transformed by taking the first differences of the natural logarithms of the values in each the series so as to attain stationarity in the first moment equation using the equation $X_t = \ln(P_t) - \ln(P_{t-1})$,where P_t represents the weekly average value for each series. The basic statistical properties of the data are presented in Table 1. The mean returns are all positive and close to zero a characteristic common in the financial return series. All the four series have very heavy tails showing a strong departure from the Gaussian assumption. The Jarque-Bera test also clearly rejects the null hypothesis of normality. Notable is the fact that all the four series exhibit positive Skewness estimate. This means that there are more observations on the right hand side.

Table 1: Basic statistical properties of returns

	NSE INDEX	NBK	BAMBURI	KENYA AIRWAYS
Mean	0.000919	0.001650	0.003070	0.002587
Median	-0.000116	0.000000	0.000395	0.000536
Maximum	0.609314	0.601761	0.604465	2.244758
Minimum	-0.602485	-0.440545	-0.608999	-2.240928
Std. Dev.	0.044259	0.086715	0.058789	0.139903
Skewness	0.208853	1.287552	0.165014	0.046429
Kurtosis	139.1245	15.20180	42.97123	220.4956
Jarque-Bera	393763.8	3900.838	40078.36	1186550
Probability	0.000000	0.000000	0.000000	0.000000
Observations	510	602	602	602

The series having exhibited heteroscedasticity as shown by the time plots were tested for the ARCH disturbances using Engle's (1982) Langrage Multiplier (LM) while the Portmanteau Q test (McLeod and Li, 1983) based on the squared residuals was used to test for the independence of the series. Since both the Q statistic and the LM are calculated from the squared residuals, they were used to identify the order of the ARCH process. For all the return series, the Q statistics and the Langrage Multiplier (LM) tests indicated strong heteroscedasticity for all the lags from 1 to 12.This suggested an ARCH model of order q=8 or a GARCH(1,1).

5.2 Empirical Results and Discussions

5.2.1 Bilinear models

The MLE method assuming a Gaussian distribution was used for parameter estimation for the data studied. Order selection was done using the ACF and PACF.

The model adequacy was checked via Ljung-Box Q statistics as well as checking the residual and squared residuals, ACF and PACF which all showed that all the residuals for bilinear models were not significantly correlated to lag 12 except the squared residual for NBK which was significantly correlated at lag 12. This implies that the fitted bilinear models were adequate except the one for NBK.

The Jarque-Bera (1980) statistics rejected the null hypothesis of normality in all the residuals i.e the residuals are not normally distributed. The fitted bilinear models and model diagnostics are presented in Appendix 1 and 2 respectively.

Table 2: Maximum Likelihood Estimates for the bilinear models

NSE 20-SHARE INDEX

$$X_t = \underset{(0.87)}{0.1582} X_{t-1} + \underset{(0.524)}{0.06912} X_{t-2} + \underset{(0.521)}{0.04628} X_{t-3} - \underset{(0.324)}{0.85108} X_{t-1}\varepsilon_{t-1} - \underset{(0.313)}{0.758197} X_{t-2}\varepsilon_{t-1} - \underset{(0.811)}{0.1049} X_{t-3}\varepsilon_{t-1} + \varepsilon_t,$$

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

NBK

$$X_t = \underset{(0.143)}{-0.3369} X_{t-1} - \underset{(0.186)}{0.0907} X_{t-2} + \underset{(0.132)}{0.0100} X_{t-3} - \underset{(0.132)}{0.3499} \varepsilon_{t-1} - \underset{(0.0728)}{0.5860} X_{t-1}\varepsilon_{t-1} - \underset{(0.01056)}{0.3607} X_{t-2}\varepsilon_{t-1} + \underset{(0.022)}{0.3878} . X_{t-3}\varepsilon_{t-1} + \varepsilon$$

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

BAMBURI CEMENT LTD

$$X_t = \underset{(0.928)}{0.00992} X_{t-1} - \underset{(()904)}{0.0115} X_{t-2} - \underset{(0.449)}{0.0323} X_{t-3} + \underset{(0.006)}{1.154} X_{t-1}\varepsilon_{t-1} + \underset{(0.625)}{0.2756} X_{t-2}\varepsilon_{t-1} - \underset{(0.818)}{0.495} X_{t-3}\varepsilon_{t-1} + \varepsilon,$$

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

KENYA AIRWAYS (KQ)

$$X_t = \underset{(0.057)}{-0.4817} X_{t-1} - \underset{(0.626)}{0.0862} X_{t-2} - \underset{(0.056)}{0.5614} \varepsilon_{t-1} - \underset{(0.695)}{0.0769} \varepsilon_{t-2} + \underset{(0.196)}{1.3866} X_{t-1}\varepsilon_{t-1} + \underset{(0.349)}{0.4657} X_{t-2}\varepsilon_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

The p-values for the parameter estimates are given in the parenthesis.
From Table 2, basing on a significance level of 0.05, it is clear that for the NBK, the observations are significant at the second lag and also the interaction between the observations and errors at lags (3,1), i.e. the bilinearity is significant at (3,1).

In case of the Bamburi series, the interactions between the observations and the errors are significant at all the lags (1,1) for the fitted model. However, for the Kenya Airways series, all the estimated parameters are not statistically significant i.e the dependence on the observations and errors are not significant. This implies that the model for the Kenya Airways is not useful and hence should be discarded but it was kept for comparison purposes with the other models.

It is worth noting that in the bilinear model, a lot of parameters have been estimated. This goes against the principle of parsimony where by models with fewer parameter estimates are preferred. The Goodness of fit statistics and the Diagnostics for the bilinear models are presented in Appendices 1 and 2 respectively.

All the fitted models had a significant correlation in the squared standardized residuals. This calls for fitting the Bilinear-GARCH (BL-GARCH) models were fitted to the respective series employing the MLE method with the Gaussian distribution assumption. The variance equation was aimed at capturing the second order correlation thereby improving the model adequacy and stability.

5.2.2. Bilinear-GARCH models

Table 3 presents the estimated BL-GARCH models for the stock under consideration.

Table 3: Estimated Bilinear-GARCH models

NSE 20-SHARE INDEX

$$X_t = \underset{(0.00)}{0.40205} X_{t-1} - \underset{(0.484)}{0.03573} X_{t-2} + \underset{(0.004)}{0.0988} X_{t-3} + \underset{(0.215)}{0.700} X_{t-1}\varepsilon_{t-1} + \underset{(0.388)}{0.3739} X_{t-2}\varepsilon_{t-1} + \underset{(0.1699)}{0.5115} X_{t-3}\varepsilon_{t-2} + \varepsilon_t,$$

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^{\,2})$$

$$h_t = 0.01314 + \underset{(0.017)}{0.6041}\varepsilon_{t-1}^2 + \underset{(0.001)}{0.3475} h_{t-1}$$

NBK

$$X_t = -\underset{(0.00)}{0.4318} X_{t-1} - \underset{(0.109)}{0.1760} X_{t-2} - \underset{(0.002)}{0.0883} X_{t-3} - \underset{(0.00)}{0.5957}\varepsilon_{t-1} - \underset{(0.027)}{0.1053} X_{t-1}\varepsilon_{t-1} + \underset{(0.001)}{0.1859} X_{t-2}\varepsilon_{t-1} + \underset{(0.00)}{0.1832} X_{t-3}\varepsilon_{t-1} + \varepsilon_t,$$

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^{\,2})$$

$$h_t = 0.0483 + \underset{(0.001)}{0.8613}\varepsilon_{t-1}^2 + \underset{(0.7)}{0.0496} h_{t-1}$$

BAMBURI

$$X_t = \underset{(0.001)}{0.1921} X_{t-1} + \underset{(0.201)}{0.07104} X_{t-2} - \underset{(0.843)}{0.0246} X_{t-3} + \underset{(0.1633)}{1.0727} X_{t-1}\varepsilon_{t-1} + \underset{(0.2024)}{1.1773} X_{t-2}\varepsilon_{t-1} + \underset{(0.103)}{6.554} X_{t-3}\varepsilon_{t-1} + \varepsilon_t,$$

$$\varepsilon_t \sim WN(0, \sigma_\varepsilon^{\,2})$$

$$h_t = 0.0208 + \underset{(0.213)}{0.7607}\varepsilon_{t-1}^2 + \underset{(0.179)}{0.2403} h_{t-1}$$

KENYA AIRWAYS (KQ)

$$X_t = \underset{(0.831)}{-0.1729} X_{t-1} - \underset{(0.967)}{0.00607} X_{t-2} - \underset{(0.657)}{0.3496}\varepsilon_{t-1} - \underset{(0.966)}{0.0104}\varepsilon_{t-2} + \underset{(0.213)}{0.9225} X_{t-1}\varepsilon_{t-1} + \underset{(0.445)}{1.2789} X_{t-2}\varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim WN(0, \sigma_\varepsilon^{\,2})$$

$$h_t = 0.0341 + \underset{(0.004)}{0.6387}\varepsilon_{t-1}^2 + \underset{(0.382)}{0.1352} h_{t-1}$$

The probability values for the parameter estimates are given in the parenthesis

The variance equation for the NSE index shows that the estimates for both β and α are significant at 5% significance level. For NBK the , α is significant while β is not significant. The Bamburi series the, parameter estimates for the conditional variance equation were all non significant at 0.05 significance level. The KQ series, α in the variance equation was significant at 5% significance level. The variance equation in all cases gives the sum of α_1 and β_1 approximately equal to 1 confirming the stability of the fitted models. The sum of the parameters α and β gives the rate at which the response function decays (Frimpong and Oteng-Abayie, 2006). This implies that the volatility in the Nairobi Stock Market is highly persistent and all the information is important in forecasting a given stock.

The fitted BL-GARCH models were diagnosed using AIC, SBC and the log likelihood ratio test. The Gaussian MLE criterion was used in parameter estimation for the BL-GARCH models. The BL-GARCH models fitted are adequate since the standardized residuals and squared residuals are not significantly correlated as shown by the Ljung-Box Q statistics. In addition, the J-B statistics strongly rejected the null hypothesis of normality in the residuals for all the series. The Goodness of fit statistics and the diagnostic checks for the bilinear models and the BL-GARCH models are presented in Appendices 1and 2 respectively.

5.2.3 Efficiency Evaluation in Bilinear and Bilinear-GARCH Models

The model efficiencies were once more evaluated using the Mean Squared Errors. The models that had the minimal MSE were considered the most efficient. However, other statistical properties especially the diagnostics and goodness of fit tests were considered in choosing the most efficient models. The MSE for the bilinear and bilinear-GARCH models are presented in Table 4

Table 4: MSE for bilinear and bilinear-GARCH models

Series	BILINEAR	BL-GARCH(1,1)
NSE INDEX	0.0006592	0.00071760
NBK	0.0064124	0.00725612
BAMBURI	0.0025673	0.00324899
KQ	0.002854	0.0028990

Despite the bilinear models having a relatively smaller MSE, they are unstable as manifested by the residual time plots and hence could be unsuitable for modelling stocks data. However, this problem is easily solved by the inclusion of GARCH models**.** This is because the GARCH model captures the heteroscedasticity properties of the series.

The Kurtosis for the BL-GARCH models are the lowest compared to the ones for bilinear models. This means that the BL-GARCH has successfully captured the heavy tail in the conditional variance of the stock market data. This could be due to interactions between past shocks and volatility in the data. The BL-GARCH is a better alternative to the bilinear models. In conclusion, the BL-GARCH models are better than the bilinear models as far as the efficiency and statistical properties (diagnostics and goodness of fit) are concerned when applied to the Nairobi Stock data.

5.2.4. Comparison between Bilinear and Bilinear-GARCH models

A comparison between the three classes of models was done based on the diagnostic test, goodness of fit statistics in addition to the MSE which showed the efficiency for each model. The bilinear-GARCH models assuming Gaussian distribution emerged as the most efficient models for modelling stock market data while the pure bilinear models were the worst in terms of model adequacy and efficiency. This is an indication that the non-linearity in the data sets are best modelled by the bilinear models while the non-stationarity are best captured by the GARCH. Thus a bilinear-GARCH model simultaneous captures the nonlinearity and heteroscedasticity.

5.2.5 Conclusions

Considering the bilinear model, the Gaussian assumption was more appropriate when employing the MLE criterion. In addition, the models seemed to have many cases of convergence problems when the MLE was implemented. The residual time plots for the bilinear models manifested sharp spikes outside the standard error band. This implies the instability of the bilinear models. Interestingly, the MSE for the bilinear emerged the lowest in all cases. This is quite challenging since the models seem very efficient but could not be considered due to their instability.

To address the instability manifested by the residuals for bilinear models, a GARCH (1,1) was fitted to the residuals of the bilinear models. MLE assuming a Gaussian distribution was utilized. The results indicated an improvement as far as the residuals are concerned especially in the reduction of residual Kurtosis. The BL-GARCH captured the asymmetry better than the bilinear.

In summary, a comparison between the two classes of models was done based on the diagnostic test, goodness of fit statistics in addition to the MSE which revealed the efficiency for each model. The bilinear-GARCH models assuming Gaussian distribution emerged as the most efficient models for modelling stock market data while the pure bilinear models were the worst in terms of model adequacy and efficiency.

6.0 References

Bera, A.K. and Higgins, M.L. (1997). ARCH and Bilinearity as Competing models for Nonlinear dependence. *Journal of business and Economic Statistics*, Vol **15** No.1, 43-50.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of econometrics*, **31**, 307-327.

Curto, J.J.D.(2002). *Modelling the volatility in the Portuguese Stock Market: A Comparative Study with German and US markets*. Unpublished Phd Thesis, Lisbon University.

Davidson, E.H.J. (2008). Time series modelling in http://www.timeseriesmodelling.com/

Degiannakis, S. and Xekalaki, E. (2004). Autoregressive Conditional Heteroscedasticity Models: A Review. *Quality Technology and Quantitative Management*, **1(2),** 271-324.

Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K inflation, *Econometrica*, **50**, 987-1008.

Engle, R.F. (1983). Estimates of the Variance of US Inflation Based on the ARCH Model. *Journal of Money Credit and Banking*, **15**, 286–301.

Engle, R.F. (2004). Risk and Volatility: Econometric Models and Financial Practice. *The*
American Economic Review, **94**, 405-420.

Granger, C.W.J and Andersen, A.P. (1978). *An introduction to bilinear time series models.* Vandenhoeck and Ruprecht. Gottingen.

Guegan, D and Pham, D. T. (1989). A note on estimation of the parameters of the diagonal bilinear model by the least squares method. *Scand. J. Stat.*,**16**, 129-136.

Maravall, A. (1983). An application of nonlinear time series forecasting. *Journal of Business and Economics Statistics* ,**1**,66-74.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) models. *Econometric Theory* **6**, 318-334.

Slustzky, H.(1937). The summation of random causes as the source of cyclic process. *Econometrica*. **5**, 105-146.

Subba, R. T. and Gabr, M.M (1984). *An introduction to Bispectral Analysis and Bilinear Time Series Models*. Lecture notes in Statistics, 24. New York: Springer.

Subba, R. T. (1981). On the theory of bilinear models. *Journal of Royal Statistics Society* ,Series B, **43**,244-255.

Subba, R. T. and Gabr, M.M (1984). *An introduction to Bispectral Analysis and Bilinear Time Series Models*. Lecture notes in Statistics, 24. New York: Springer.

Tong, H. (1990). *Non- Linear Time Series. A Dynamical System Approach*. Oxford Science Publications.

Wagala. A., Nassiuma, D.K, Ali, S. I and Mwangi, J. W (2012). Volatility Modelling Of
The Weekly Returns Of The Nairobi Securities Exchange Using The ARCH-Type Models. *International Journal of Applied Science and Technology, Vol. 2. No. 3, 165-174*

Weiss, A.A (1986). ARCH and Bilinear Time Series Models: Comparison and Combination. *Journal of business and Economic Statistics,* Vol **15** No.1, 59-70.

Appendices

Appendix 1: Goodness of fit statistics for the bilinear models

		BILINEAR (GAUSSIAN)	BL-GARCH (GAUSSIAN)
NSE INDEX	LR	1462.65	1647.89
	AIC	-1454.65	-1636.89
	SBC	-1436.73	-1612.25
NBK	LR	822.73	953.732
	AIC	-813.73	-917.366
	SBC	-792.976	-942.732
BAMBURI	LR	1163.24	1322.72
	AIC	-1156.24	-1313.72
	SBC	-1140.1	-1292.96
KQ	LR	1125.42	1237.34
	AIC	-1118.42	-1228.34
	SBC	-1102.27	-1207.58

Appendix 2: Diagnostic Tests for the bilinear models

Series	Statistics	BILINEAR	BL-GARCH(1,1)
NSE INDEX	Skewness	0.22056	0.6527
	Kurtosis	8.9535	10.2272
	JB	962.986 (0.00)	1465.27 (0.000)
	Q(12)	18.3504 (0.59)	10.6459 (0.386)
	$Q^2(12)$	112.201 (0.000)	6.7996 (0.912)
NBK	Skewness	1.1355	0.0883
	Kurtosis	15.2499	10.1113
	JB	4811.76 (0.00)	1568.64 (0.00)
	Q(12)	15.6521 (0.208)	14.7397 (0.256)
	$Q^2(12)$	48.0631 (0.00)	3.1931 (1)
BAMBURI	Skewness	-2.273	0.5185
	Kurtosis	38.0157	13.313
	JB	38649.7 (0.000)	3330.41 (0.00)
	Q(12)	19.5401 (0.107)	19.3608 (0.112)
	$Q^2(12)$	17.6902 (0.342)	7.7882 (0.955)
KQ	Skewness	0.4617	0.4954
	Kurtosis	9.2534	8.4799
	JB	1240.35 (0.00)	962.624 (0.000)
	Q(12)	15.965 (0.193)	19.4339 (0.079)
	$Q^2(12)$	154.448 (0.000)	7.741 (0.956)